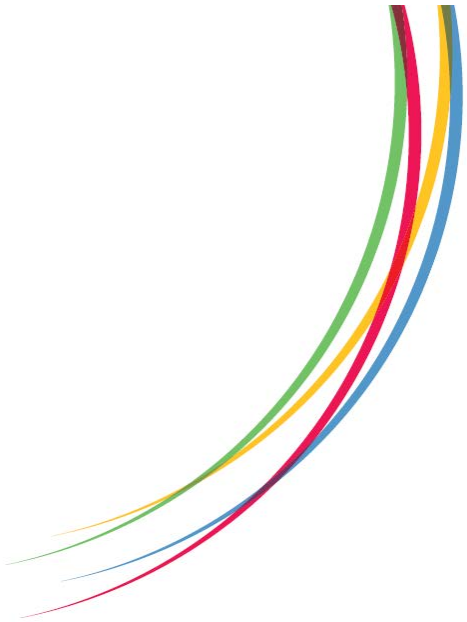

UNDERSTANDING SOCIETY: UK HOUSEHOLD LONGITUDINAL STUDY: USER GUIDE TO ETHNICITY RESEARCH

Stephanie McFall
Alita Nandi
Lucinda Platt

Institute for Social and Economic Research
University of Essex
Colchester
Essex

Version 1, December 2014

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

CONTENTS

Understanding Society: UK Household Longitudinal Study: User Guide to ethnicity research 1
1. Overview of the study 3
2.Studying ethnicity in a longitudinal household panel survey 4
3.Sample design for surveying ethnic minorities 6
3.1 Understanding society sample design 6
3.2 Ethnic Minority Boost sample 7
3.3 Sample status and following rules 10
4.Specific content 10
4.1 Extra Five Minutes questions 10
How to identify the extra five minutes questions? 11
4.2 The Extra Five Minutes sample 12
The Ethnic Minority Boost sample 12
The General Population Comparison sample 12
The Ethnic Minority in Low Ethnic Minority Density Area sample 12
5.Response and attrition 13
6.Data files and question coverage 16
6.1 Data structure, naming conventions 16
6.2 Key variables for ethnicity research 17
7.Issues for analysis: sample design, non-response, weighting 19
7.1 Why use weights? 19
7.2 Available weights 21
7.3 Sample design variables and analysis 22
8.Approach to translation and language 22
9.Help and support 24
10.Citation information and credits 24
Citation of this Guide 25
11.Additional links 25
12.References 26
13.Ethnicity and migration research using Understanding Society 27
Journal articles and Working papers 27
Book chapters and reports 27

1. OVERVIEW OF THE STUDY

This guide provides an introduction to carrying out ethnicity related research with *Understanding Society*. It is intended to complement other documentation providing support and guidance on the study.

Understanding Society -- the UK Household Longitudinal Study (UKHLS) -- is a large longitudinal survey of households in the United Kingdom (England, Scotland, Wales and Northern Ireland). Members of households recruited at the first round of data collection are visited one year later to collect information on changes to their household and individual circumstances. *Understanding* Society is funded by the Economic and Social Research Council with additional support from multiple government departments. The scientific leadership team is from the Institute for Social and Economic Research (ISER) of the University of Essex, the University of Warwick, and the London School of Economics and Political Science.

Each year, since the survey started in 2009, all eligible adult respondents in participating households are interviewed. They also complete a self-completion questionnaire. Children aged 10-15 also complete a separate self-completion questionnaire, becoming eligible for the main questionnaire when they reach 16. Over time a detailed picture of individuals' lives builds up through the annual interviews. In addition, each annual survey provides information about changes and continuity in respondents' lives since the last interview, changes which can be related to their characteristics and to other events that have taken place in their lives.

Data from each wave of the study is available from the UK Data Service (see Data Access, below). Analysts should also consult the User Manual for further infomration on the overall study (Knies 2014).

The purpose of *Understanding Society* is to provide high quality longitudinal data about subjects such as health, work, education, income, family, and social life to help understand the long term effects of social and economic change, as well as policy interventions designed to impact upon the general well-being of the UK population.

Research on ethnicity and studies looking across the UK's ethnic groups was one of the original aims of *Understanding Society*. This was reflected in three core aspects of the study. First, the main sample was supplemented by an ethnic minority boost sample of over 4,000 households. Along with the ethnic minorities in the main sample, this boost sample enables detailed analysis of and comparison across individual ethnic groups. We discuss the design and implications for analysis of the inclusion of the boost sample further below.

Second, qustionnaire content that was salient for research into ethnicity and comparisons across ethnic groups was collected in the main questionnaire, subject to an extensive consultation process. For example, questions on country of birth of the respondent, their parents and grandparents were collected of all respondents in the first survey; and questions on health and wellbeing, which are of great interest to ethnicity researchers as well as to others, are also collected of all respondents in annual interviews.

Third, an additional five minutes of question time was allocated to a sub-sample of respondents, in order to capture questions of specific interest for ethnicity research, in particular research relating to the experience of ethnic minorities in the UK. The coverage and implications for analysis of the extra five minutes sample are discussed further below.

2. STUDYING ETHNICITY IN A LONGITUDINAL HOUSEHOLD PANEL SURVEY

Understanding Society brings an unprecedented opportunity to study issues of continuity and change within and across the UK's ethnic groups. It will enable analysis of inequalities, and of transitions over time are different or similar or lead to different or similar outcomes for those of differente ethnic backgrounds. It will enable policy makers and researchers to understand better the complexity of the current population and to identify those factors that drive inequalities.

Nevertheless, studying ethnicity in survey research brings a number of challenges. The concept of ethnicity is contested and its definition and operationalisation varies across disciplines. Moreover there are distinctions in the extent to which researchers and analysts adopt essentialist attitudes to ethnicity or regard ethnic groups as proxying for other factors, or collections of factors, that are (currently) unmeasured. See further, the discussion in Burton et al. 2010.

The concept of ethnicity can be approached in a number of ways. For example, it can be understood as commonality within a group or as differences from 'other' groups. Its multifaceted nature makes measurement using a single measure challenging, and yet that is often what is available to researchers. Researchers bring their own agendas to the topic: while social psychologists may be interested in understanding individuals' identity, sociologists may be interested in social stratification and the role of ethnicity within that. Other social and health scientists may be primarily interested in the role of immigration in shaping individuals outcomes and trajectories in the country of destination

As a result, *Understanding Society* incorporates an exceptionally rich and varied range of measures to capture different dimensions of ethnicity and ethnic group identity, namely:

1) *ONS Census, single category questions on ethnic group (See further, the detail on the ONS page)*

Developing an ethnic self-identification (categorical) question is a major challenge. In addition to standard survey design issues such as question wording, the design of such a question requires a fine balancing act between trying to get consistent, reliable measures of ethnic identity and capturing people's perception of their own ethnic identity. The response categories need to be meaningful, acceptable and consistent, so that people will be able to respond to them in expected, and predictable, ways. The ONS question represents a national-level attempt to provide such a simple single-category solution. While it is subject to some level of critique, it is widely used across government and other surveys as well as administrative sources, and hence provides an important point of reference for and comparison. See further the documents relating to the recommended questions here.

2) National identity, multiple response

As with the implementation of the ONS ethnic group question in the 2011 Census, *Understanding Society* includes the multiple response national identity question that was asked directly before the ethnic group question. The inclusion and positioning of the national identity question in the 2011 Census was designed to enable respondents to express national identity separately from ethnic group and hence, ideally, to provide more responses that did not confound national identity with ethnic identity at the ethnic group questions. Again this is asked of all Understanding So

3) Ethnic minority boost screening question

The composition of the ethnic minority boost (EMB) sample for *Understanding Society* was intended to maximise the analytical potential of the sample for comparisons of ethnic groups and within minority groups. The screen question was therefore intended to capture a number of specific groups that have some internal coherence and are sufficiently numerous for meaningful analysis. Thus its operationalization was driven by pragmatic rather than analytic concerns. It is not intended to reflect ethnic identity, but is nevertheless available for researchers if they wish to understand how respondent households came to be included in the sample.

4) Country of birth (across multiple generations), and parental ethnic group

Immigrant status is of core interest to a wide range of researchers. Thus it was vital to include respondent's country of birth – and time of arrival in the UK. However, immigrant *generation* is also of substantial interest, hence the country of birth of the respondent's parents and grandparents were also asked for the whole sample. This not only allows outcomes across immigrant generation to be tracked, it also enables researchers to relate ethnic identity to associations with countries outside the UK or to construct measures of ethnicity that combine information from across these different sources on heritage. Moreover they can also relate measures of respondent's ethnic group or country of birth or parental country of birth to measures of parents' (father's and – if different – mother's) ethnic group that are also included.

5) Strength of British identity and of identification with parental ethnicity

There is an extensive and developing literature on identification of minority groups with both majority and minority identities, and how this relates to constructs of ethnic group and of relationship to the majority society. For those interested in such issues, there are measures of strength of British identity and strength of identification with parents' ethnicity.

6) Measures of identity, belonging to and pride in domains of ethnic identity

To better understand how identity is expressed and experienced at different points in the life course and in relation to ethnic a suite of questions are included in *Understanding Society* that tap not only into ethnic group, which may potentially primarily affect ascribed rather than 'owned' ethnicity, a series of questions tap into different psychological dimensions of identity (pride, belonging etc.) across a series of domains (country of birth, ethnic group etc.).

7) Ethnicity of associates

Social understandings of ethnicity and its behavioural and social conception or implications are assisted by measures that identify ethnicity of friends and networks.

Such measures, that also identify other characteristics of friendship networks, are therefore available.

8) Harassment and discrimination

The salience of ethnicity and ethnic group identity has been linked to the boundaries that are set up to exclude those regarded as 'other'. Hence measurement of experience of (ethnically-based) discrimination or harassment can provide insight into how

9) Religion and language

Measures that may be associated with particular ethnic or national origins, or be used in construction with them are also of value in enabling researchers to operationalise ethnic group constructs that are of most interest to them. Of these religion, enabling the construction of ethno-religious groups, is a key measure that *Understanding Society*; language is also embedded in constructions of ethnicity and minority status.

Some additional information on these variables, including variable names and when asked, is provided further, below

Overall, *Understanding Society* furnishes researchers with a large repertoire of ethnicity related questions to enable them to carry out analysis according to the framework of their discipline and research questions. These have been discussed in relation to how they are linked to construction of ethnicity / ethnic group itself. Clearly many measures (such as harassment or identity) can also be investigated to the extent that they are associated with (different) ethnic groups, along with the other rich set of multi-topic measures carried across the study: full questionnaires for each of the waves to date are available on the *Understanding Society* web pages.

3. SAMPLE DESIGN FOR SURVEYING ETHNIC MINORITIES

3.1 UNDERSTANDING SOCIETY SAMPLE DESIGN

Understanding the sample design is important for any research using Understanding Society data including ethnicity research. The *Understanding Society* sample has multiple components. The design of all components is described in more detail in two *Understanding Society* working papers (Lynn 2009, Berthoud et al 2009).

- The **General Population Sample** consists of two separate samples of residential addresses: one for Great Britain (that is, England, Scotland and Wales) and another one for Northern Ireland. The **Great Britain sample** is a proportionately stratified (equal probability), clustered sample of addresses selected from the Postcode Address File. The **Northern Ireland sample** has an unclustered systematic random sample of addresses selected from the Land and Property Services Agency list of domestic addresses, with selection probability which is twice from that of the Great Britain sample. Interviewing for the main sample takes place over 24 months, except for Northern Ireland where it takes place over 12 months. Each monthly sample of the GB part of

the GP sample is a random sub-sample but that is not the case for the other samples. At wave 1 the GPS had a sample size of approximately 26,000 participating households.

- The **former BHPS sample** became a part of the UKHLS at Wave 2. It has all members from the BHPS sample who were still active at Wave 18 of the BHPS and who had not refused consent to be issued as part of the *Understanding Society* sample. It should be noted that the BHPS sample itself has multiple components: a nationally representative sample of addresses in Great Britain (south of the Caledonian Canal) in 1990, boost samples of Scotland and Wales added in 1999,and a boost sample of Northern Ireland addresses added in 2001. For further details of the BHPS sample, see section IV of the BHPS User Guide. The BHPS sample data collection takes place over 12 months for each wave. At wave 2 of the UKHLS, the BHPS sample had a sample size of approximately 6,600 participating households.

- **Ethnic minority boost sample** (EMB), like the main sample, is collected over a 24 month surver period for each wave. ***The design is described further below***. At wave one the EMB comprised approximately 4,000 participating households.

For all sample components, once addresses were selected, up to 3 dwelling units at each address were randomly selected and at each dwelling unit upto 3 households were randomly selected. Dwelling units are defined in interviewer instructions as "a living space with its own front door – this can be either a street door or a door within a house or block of flats. Usually there is only one dwelling unit at an address", while households are defined as the person or people occupying particular shared spaces, namely, "a **household** is when one person or a group of people have the accommodation as their only or main residence **and** share at least one meal a day **or** share the living accommodation" Around 95% of addresses have one DU and one household - multiple DUs/households tend to be concentrated in urban areas.

3.2 Ethnic Minority Boost sample

The Ethnic Minority Boost sample was designed to provide at least 1,000 adult interviews from each of five target ethnic minority groups: Indian, Pakistan, Bangladesh, Caribbean, and African, as well as somewhat less comprehensive coverage of other minority ethnic groups within the sampled areas.

The EMB sample design comprised of screening addresses in high ethnic minority concentration areas in Great Britain. **Please note that Northern Ireland does not contribute to the EMB sample.** The initial step was identifying postal sectors with relatively high proportions of relevant ethnic minority groups, based on 2001 Census data and more recent Annual Population Survey data. This identified 3,145 sectors or approximately 35% of the sectors in Great Britain. It covers between 82% and 93% of the population of the five ethnic minority groups in the UK.

The 3,145 sectors were sorted into four strata based on the expected number of ethnic minority households that would be identified by the sampling and screening procedures (see Berthoud et al., 2009 for details). All sectors were included for the stratum where a yield of three or more households was expected. In the other three strata, sectors were sub-sampled at rates of 1 in 4, 1 in 8, or 1 in 16 respectively.

This was done to constrain the number of sectors that might have just one or two eligible sample households (or even none). The total number of postal sectors selected for inclusion in the Ethnic Minority Boost sample was 771. Of these 6 were in Scotland, 7 were in Wales, and the remaining 758 were in England, with a concentration in London (412 sectors). Note that the geographic concentration of ethnic minority populations influences the distribution of selected postal sectors from which the EMB was sampled.

The number of addresses selected per postal sector ranged from 15 to 103. Sampling fractions varied across the sectors in a way designed to deliver target numbers of respondents in each target ethnic minority group with adequate statistical efficiency (see Berthoud et al., 2009 for more details). In sectors selected for both the General Population Sample component and the Ethnic Minority Boost sample, a single systematic sample of the required total number of addresses was selected, so there was allocation to spread the EMB and GPS throughout the whole sector.

Once the dwelling unit and household was selected, the final stage of sampling was screening conducted by the interviewers. The Understanding Society Wave 1 Project Instructions for interviewers outlines the procedures. As a result the Address Record Forms for the EMB sample were different from those used with the General Population Sample. Within each household in each selected address, rather than all resident persons becoming sample members, there were two additional steps:

1) A “screen” was carried out to identify whether there were any persons who were from any of the specified minority ethnic groups in the household. The household member answering the screening question was asked **“Does anyone living at this address come from, or have parents or grandparents from any of the following ethnic groups?”** The ethnic groups that they could choose from were Indian, Mixed Indian, Pakistani, Bangladeshi, Sri Lankan, Caribbean/West Indian, Mixed Caribbean/West Indian, North African, Black African, African Asian, Chinese, Far Eastern, Turkish, Middle Eastern / Iranian AND None of these.

2) If the response was anything other than “None of these” then the household had a positive selection probability (“sampling fraction”). At the outset the sampling fraction was not 100 per cent for all ethnic groups, but differed across ethnic groups according to the sample design stage (see Berthoud et al 2009). In other words, not every household that included someone with an ethnic minority background was chosen. A review after six months concluded that sampling fractions for all groups had to be increased and additional addresses had to be screened for Bangladeshi members to reach the target. As a result, for most groups the sampling fractions were increased to 100 per cent. The only exception was Indians who would otherwise have been sampled in sufficiently large numbers that they would have dominated the overall boost sample, and hence undermined attainment of the target sizes for the other target groups. Table 1a shows the selection fractions by ethnic groups.

The distribution of ethnic groups reported by households in response to the “screener” is shown in Table 1b. As some households include members from

different ethnic groups, these numbers will not add up to all the households in the EMB sample (screened in).

Table 1a: Secondary screening retention fractions

Main category	Sub-group	Fraction for assignments issued in 2009	Fractions for assignments issued in 2010
Indian	Indian (non-mixed)	50%	65%
	Mixed Indian	100%	100%
Pakistani	Pakistani	54%	100%
Bangladeshi	Bangladeshi	100%	100%
Caribbean	Caribbean (non-mixed)	91%	100%
	Mixed Caribbean	100%	100%
African	African	82%	100%
Other included groups	Sri Lankan	100%	100%
	Chinese	100%	100%
	Turkish	100%	100%
	Far eastern	30%	100%
	Middle eastern	30%	100%

Source: Berthoud et al 2009

Table 1b: Number of households reporting at least one household member (or their parents or grandparents) come from different ethnic group[a]

Indian	2303
Mixed Indian	234
Pakistani	1645
Bangladeshi	955
Sri Lankan	265
Caribbean/West Indian	1785
Mixed Caribbean/West Indian	491
North African	187
Black African	1730
African Asian	93
Chinese	451
Far Eastern	402
Turkish	224
Middle Eastern Iranian	463

[a] Note some households reported members from more than one ethnic group and so these numbers will not add up to all the households in the EMB sample (screened in)

Once selected, the sample was treated as per the General Population Sample.

The overall sampling fractions combine a) the probability of sampling the sector, b) the fraction of addresses selected within the sector, and c) the probability of a

household being retained following the application of the random selection mechanism described above.

3.3 SAMPLE STATUS AND FOLLOWING RULES

Following rules determine whether specific household members should be asked to be interviewed in the next wave. The purpose of these following rules is to make sure that the study continues to be representative of the original sample selected in 2009/10 and their descendents and information on the household context is always collected. For ease of implementation of these following rules, all household members are classified into three sample statuses: Original Sample Members (OSMs), Temporary Sample Members (TSMs), and Permanent Sample members (PSMs).

OSMs and PSMs, of all ages, are followed and remain eligible for interview as long as they are resident within the UK, potentially for the life of the survey. TSMs remain eligible for interview as long as they are co-resident with an OSM/PSM (including when the only OSM in the household is a child who is not yet eligible for personal interview). TSMs who are not co-resident with an OSM/PSM are not followed and become ineligible for interview.

OSM: All members of *Understanding Society* General Population Sample households enumerated at Wave 1, including absent household members and those living in institutions who would otherwise be resident, are OSMs. All **ethnic minority members** of an enumerated household eligible for inclusion in the Ethnic Minority Boost sample are OSMs. In all of these samples, any child born to an OSM mother after Wave 1 and observed to be co-resident with the mother at the survey wave following the child's birth is an OSM.

TSM: Anyone co-resident with a OSM/PSM from onwards 2 is considered to be a TSMs.). This would include any child born to an OSM father after Wave 1 but not an OSM mother and observed to be co-resident with the father (or any other OSM) at the survey wave following the child's birth. Members of an enumerated household eligible for inclusion in the Ethnic Minority Boost sample at Wave 1 who are not from a qualifying **ethnic minority group** are also considered to be TSMs (this was the only category of TSM at Wave 1.

PSM: Any TSM father of an OSM child born after Wave 1 and observed to be co-resident with the child at the survey wave following the child's birth is considered to be a PSM.

4. SPECIFIC CONTENT

4.1 EXTRA FIVE MINUTES QUESTIONS

Given the importance of the ethnicity strand of *Understanding Society* to the study as a whole, and the potential for addressing specific research interests in a way that had not been possible since the Fourth National Survey of Ethnic Minorities in 1993/4, it was determined from the outset to collect an extra five minutes' worth of questions for specific topics of interest for ethnicity research. Of particular priority were those areas which would enable longitudinal research on pertinent topics.

However, extra questions add to respondent burden and overall survey costs and specialist questions could not be fielded across the whole sample, hence it was decided that these would be asked of a sub-sample which would allow comparative analysis. We refer to this as the "Extra Five Minute" sample, which is described further below.

The topics to be carried and addressed to the "Extra Five Minutes" sample, and their rotation over waves of the study was subject to extensive consultation with the ethnicity researchers, co-funders, third sector and the Ethnicity Strand Advisory Committee. As a result of this consultation, the broad areas for inclusion and their rotation were determined as illustrated in the top panel Table 2. There are aslo som questions asked on a different (more frequent) rotation for the Extra Five Minutes sample or commencing at a different starting point due to particular interests or topics within them. These are illustrated in the second panel of Table 2.

Table 2a: Extra Five Minute Questions

Module	Sample asked	Rotation	Wave
Migration history		Initial conditions	1
Harassment		2 year rotation	1, 3, 5, 7
Discrimination		2 year rotation	1, 3, 5, 7
Remittances		3 year rotation	1, 4, 7
Ethnic identity		6 year rotation	2, 8
Service Use			4
Religious practice		4 year rotation	4, 8

Table 2b: Modules asked of the whole sample and the Extra Five Minutes sample but on a different rotation

Module	Sample asked	Rotation	Wave
Religious belonging[a]	Whole sample	3 year rotation	1, 4
	New entrants of Extra Five Minute sample & NI residents		2,3,5,6
	All	4 year rotation	8
Political engagement	Extra Five Minute sample		2
	Non-Extra Five Minute sample		3
	All	4 year rotation	6
Britishness	Extra Five Minute sample		1,3
	All (self-completion)		6

Notes: [a]From onwards wave 8 these questions would be asked of the whole adult sample and would be on 4 year rotation

Additionally there are some modules in which most of the questions are asked of the whole sample and a few extra questions in that module are only asked of the Extra Five Minutes sample. For example, the Best Friends module asked in Wave 3 self-completion questionnaire.

HOW TO IDENTIFY THE EXTRA FIVE MINUTES QUESTIONS?

In the questionnaires these questions can be identified by searching for this description of the "Universe" (refers to those who are eligible for a particular question):

If (HHGRID.EMBoost = 1 | HHGRID.GPCompare = 1 | (HHGRID.LDA = 1 & ETHNICITYANDNATIONALIDENTITY.RACEL > 4 & ETHNICITYANDNATIONALIDENTITY.RACEL < 98)) // Ethnic Minority Boost or General population comparison sample or LDA and any non-white background

The text identifies the different components of the overall sample who receive the questions as we go on to discuss next.

4.2 The Extra Five Minutes sample

The Extra Five Minute Sample consists of three components: the Ethnic Minority Boost (EMB) sample, the General Population Comparison sample, the Ethnic minority in Low Density Area sample. In addition many of the questions, in particular those where responses might be expected to change most quickly among more recent immigrants, are asked of all recent immigrants, that is those who immigranted within the three years prior to the beginning of wave 1 ofUnderstanding Society, even if they are not in the EMB.

The Ethnic Minority Boost sample

The EMB sample includes sufficient sample sizes of different ethnic groups allows analysis separately by ethnic groups.

The General Population Comparison sample

To allow comparative anlaysis with the White majority (defined as those who choose White – British/English/Scottish/Welsh/Northern Irish as their ethnic group) group the Extra Five Minutes questions were also asked of a random sub-sample of the GP Sample, referred to as the General Population Comparison (GPC) sample. The GPC includes one sampled address from 40 per cent of the selected postal sectors in the Great Britain component GPS. In other words, of the 2,640 general population sectors, 60 per cent of them (1,584) contain 18 GPS addresses and the other 40 per cent contain 17 GPS addresses and one GPC address. The persons in these households were designated as members of the GPC sample, *regardless of ethnic group membership*.

The Ethnic Minority in Low Ethnic Minority Density Area sample

Existing studies have shown impact of the composition and quality of residential neighborhoods on various life outcomes. The EMB sample was drawn from high ethnic minority concentration areas and, as a result, any anlaysis based on the EMB and GPC samples would not be representative of the experience of all ethnic minorities. It was therefore decided to include ethnic minorities from the GP sample living in areas that would not have been eligible for inclusion in the wave 1boost screening areas on the basis that they did not have sufficient concentraitons of the target ethnic minority groups. These areas are referred to as ‘low density areas’; and hence this sample is referred to as EM-LDA sample. This status became fixed at the individual level after wave 1.

All OSM members of EMB and GPC samples and anyone co-resident with them are eligible for the extra five minute questions at each wave. At wave 1, only the EM-LDA sample members were eligible for the extra five minute questions but from wave 2 onwards anyone co-resident with them were also eligible.

5. RESPONSE AND ATTRITION

The Wave 1 survey fieldwork started on 8th January 2009 and ended on the 7th March 2011 (including the re-issue period). In total, interviews were achieved in 30,169 households (26,089 in the General Population Sample, 4,080 in the Ethnic Minority Boost sample), full or proxy interviews with 50,994 individuals (43,674 in the General Population Sample and 7,320 in the Ethnic Minority Boost sample).

Tables 3a and 3b below present the household response rates for Wave one. The individual response rates are for responding households only. In wave 1 household response rates are around 60% for the GP samples. Refusal rates are much higher in wave 1 than non-contact rates. The relatively lower household response rates of EMB sample is due to relatively higher non-contacts and refusal rates. But as these are non-responding households their eligibility for the EMB sample is not known. In other words, some of these households are possibly ineligible for the EMB sample. So, after adjusting for that the household response rate for the EMB sample will be higher. Lynn et al 2012 the adjusted EMB sample response rate to be around 52%.

Table 3a: Household response rates among eligible households, wave 1

	General Population Sample			Ethnic Minority Boost[a]
	Great Britain	Northern Ireland	Total	
Responded	57.1%	60.9%	57.3%	39.9%
Non-contact	8.6%	11.0%	8.3%	27.7%
Refused	33.9%	27.4%	33.6%	29.0%
Other	0.4%	0.6%	0.4%	3.4%
Total Eligible	43,267	2,107	45,374	10,077
Table 3b: Eligible and ineligible households, wave 1				
Eligible	89.9%	88.0%	89.8%	22.5%
Ineligible	10.1%	12.0%	10.2%	77.5%
Total Issued	48,144	2,395	50,539	44,769

[a]The non-contacts for EMB sample include those who are ineligible. So, after correcting for that response rates for the EMB sample would be higher.

Table 4 shows individual response rates (among responding households) for the different samples. Individual response rates (including proxy interviews) for the GP-GB sample is higher than that for the GP-NI and EMB samples which are similar at 80%. Refusal and non-contact rates for these two samples are also similar.

Table 4: Individual response rate (among responding households)

	General population sample			Ethnic minority boost sample
	Great Britain	Northern Ireland	Total	
Adult individual interview rate	82.0%	77.3%	81.8%	72.4%
+ proxy interview rate	87.3%	80.8%	87.0%	79.24%
Refusal rate	8.0%	11.0%	8.2%	12.2%
Non-contact rate	4.7%	8.2%	4.8%	8.6%
All enumerated adults	47,615	2,584	50,199	9,237
Youth response rate	77.2%	73.1%	77.0%	62.9%

Youth refusal rate	22.8%	26.9%	23.0%	37.1%
All enumerated youths	4,900	290	5,190	1,437
All enumerated children	8,082	477	8,559	2,687
All enumerated individuals	60,597	3,351	63,948	13,361

Table 5 shows the number of adult wave one interviews achieved by ethnic group and sample (including extra five minute sample).

Table 5: Adult wave one interviews by ethnic group and sample

	EMB sample	GP Sample	Total	Extra five minute sample
African	925	480	1405	987
Caribbean	770	349	1,119	806
Bangladeshi	950	176	1,126	967
Indian	1079	818	1,897	1,210
Pakistani	940	495	1,435	977
Five target ethnic groups	***4664***	***2318***	***6,982***	***4,947***
Arab	89	83	172	113
Chinese	191	127	318	239
Mixed	417	405	822	536
White Irish	22	695	720	25
Any other white background	125	1,253	1,378	156
Other ethnic groups[a]	653	755	1,408	888
All ethnic minority groups	***6,161***	***5,639***	***6,982***	***1,957***
White majority[b]	513	35,368	35,881	1,246
Total	***6674***	***41,007***	***47,681***	***8,150***

[a]Includes those who reported their ethnic group as Roma or Irish Traveller, Other Asian Backgroud, Other Black Background, Any Other Ethnic group. [b]Defined as White: British/ English/ Scottish/Welsh/ Northern Irish

Finally Tables 6a-6c show the response outcomes among wave one adult interview respondents across all four waves by ethnic group, religion and country of birth. Cell sizes smaller than 50 are not shown to avoid disclosure. These tables inform users of the expected sample sizes for their analyses based on ethnicity related issues where ethnicity may be measured by Census ethnic groups, religion and/or country of birth.

Table 6a: Response outcomes among wave one adult interview respondents across 4 waves by ethnic group

1: full interview, 0: non-interview including proxy interview; Position indicates wave number
For example, 1110 indicates full-adult interview in wave 1, 2 and 3 but not in in wave 4

	1000	1001	1010	1011	1100	1101	1110	1111
White British/ English/ Scottish/ Welsh/ Northern Irish	6471	215	384	985	3909	925	2819	20212
White Irish	139	*	*	*	94	*	78	358
Any other white background	440	*	*	*	198	*	109	592
Mixed	222	*	*	*	94	*	70	395
Indian	523	*	*	80	248	71	150	775
Pakistani	371	*	*	70	171	71	148	550
Bangladeshi	356	*	*	57	151	64	96	341
Chinese	115	*	*	*	52	*	*	111
Caribbean	302	*	*	*	131	*	107	497
African	446	*	*	61	236	*	142	469
Arab	94	*	*	*	*	*	*	87
Other ethnic groups[a]	293	*	*	*	132	*	85	396

[a]Includes those who reported their ethnic group as Roma or Irish Traveller, Other Asian Backgroud, Other Black Background, Any Other Ethnic group
*Denotes number of observations is less than 50

Table 6b: Response outcomes among wave one adult interview respondents across 4 waves by religion (brought up in or belongs to)

1: full interview, 0: non-interview including proxy interview; Position indicates wave number. For example, 1110 indicates full-adult interview in wave 1, 2 and 3 but not in in wave 4

	1000	1001	1010	1011	1100	1101	1110	1111
None	1602	65	115	228	864	208	599	3322
Catholic	1448	51	89	187	771	186	582	3420
Protestant	4744	140	272	754	2924	713	2057	15262
Jewish	50	*	*	*	*	*	*	93
Muslim	1173	71	115	176	537	178	374	1373
Hindu	333	*	*	*	146	*	98	453
Sikh	146	*	*	*	65	*	*	216
Buddhist	87	*	*	*	*	*	*	124
Other	177	*	*	*	92	*	61	514

*Denotes number of observations is less than 50

Table 6c: Response outcomes among wave one adult interview respondents across 4 waves by country of birth

	Non-response in some waves 2-4	Full response in waves 1 to 5
England	13900	17407
Scotland	1709	1702
Wales	922	1127
North Ireland	848	1081
Republic of Ireland	167	150
France	57	*
Germany	119	125
Italy	*	*
Spain	*	*
Poland	176	125
Cyprus	*	*
Turkey	89	*
Australia	*	*
New Zealand	*	*
Canada	*	*
USA	66	61
China/Hong Kong	147	74
India	655	441
Pakistan	551	309
Bangladesh	557	241
Sri Lanka	144	103
Kenya	113	81
Ghana	130	86
Nigeria	196	118
Uganda	56	*
South Africa	95	76

Jamaica	228	159
Other country	1801	1073

*Denotes number of observations is less than 50

While response is somewhat lower within the EMB than across the sample as a whole, both at wave 1 and over time, analysis of rates of attrition across non-UK born suggests that it is largely the characteristics of the EMB other than their ethnic group that increase the risks across the sample overall – for example, being younger, being more mobile, being more likely to live in London – that are resulting in this differential attrition.

6. DATA FILES AND QUESTION COVERAGE

6.1 DATA STRUCTURE, NAMING CONVENTIONS

We recommend the *Understanding Society User Guide* (Knies 2014) for discussion of data structure, variable names etc. Here we give a brief overview of the main elements and some key variables.

Interviewers first attempt to get some basic information such as age, sex, main activity status about all household members (enumeration grid). Next they attempt to conduct a household interview (using the household questionnaire) with an adult who is informed about the household which is genrally the person who owns or rents the accommodation or if more than one such person then the eldest of them. Next the interviewer attempts to interview all adults face-to-face (defined as being 16 years or older) in the household. All adults are asked the same set of questions but those who are members of the extra five minute sample are asked some additional questions (see Section). If someone is not available and gives permission for someone else in the household such as their partner or adult children to answer on their behalf, then that person is asked a shorter individual questionnaire which includes some factual questions but not attitudinal or subjective ones. Those who complete an adult face-to-face interview are asked to complete a self-completion questionnaire. Young persons, that is those between the ages of 10-15 are asked to complete the youth self-completion questionnaire. Children below age 10 are not directly interviewed although some information about them is collected from time to time from their parents or responsible adults. Since the second wave the surviving BHPS sample were incorporated into *Understanding Society*. Around 500 BHPS households had opted for telephone interviews and they continue to be interviewed by telephone in Understanding Society. The telphone adult questionnaire is the same as the face-to-face adult questionnaire.

As Understanding Society is a household panel survey, providing all the data in a single flat file is not very effcient. Instead different sets of files are provided for each wave. At each wave, data from adult individual interviews are stored in one file, while responses from household questionnaire are stored in another file. Details of all files provided at each wave are shown in Table 7.

The naming convention followed in Understanding Society is that all wave-specific variables and data files have a wave specific prefix: a signifies wave one, b signifies wave two and so on. While most variables and data files are wave specific there are

some which are not. Such variables and data files do not have the wave prefix. For example the cross-wave unique identifier, pidp has no wave prefix, but the main activity status variable at each one is named a_jbstat, b_jbstat at wave two and so on. Similarly the data file that stores information from adult interviews in wave one is named a_indresp, while the data file that stores time-invariant variables such as country of birth, ethnic group, sex, is named xwavedat. Also note the suffix _dv signifies the variable was derived by data managers.

Table 7: File names and content

File name	File content
w_indresp	Data from adult interviews based on individual questionnaire including proxy, telephone and self-completion questionnaires
w_indall	Enumeration and household grid information on all household members includign children and non-responding adults
w_hhresp	Data from household interviews based on household questionnaire
xwavedat	Basic time-invariant variables
w_youth	Data from youth interviews based on youth self-completion questionnaires

6.2 KEY VARIABLES FOR ETHNICITY RESEARCH

As noted above, researchers bring to the study of ethnicity different conceptual and operational frameworks for their investigations. Some focus on immigrant status (or country of birth), others are concerned with self-ascribed ethnicity, or in the construction of 'ethno-religious groups', and others are interested in ethnic identity as a subject of research in its own right. In *Understanding* Society, the question content aimed to facilitate maximum flexibility for researchers in implementing their own understandings of ethnicity and ethnic group.

Here we briefly describe the key variables available for this purpose. We provide a brief description of the question, its variable name, the sample covered and the wave (first) asked in (see Table 8). Further information on the rotation of content is provided in Table 2, above. Most questions were asked either in wave 1 or at the first time the respondent was interviewed. However, additional questions on ethnic identity were introduced in wave two and are repeated every three years for new immigrants and 16-19 year olds and every six years for everyone else.

Table 8: Key variables for ethnic group measurement and ethnicity analysis

Question	Variable Name	Description	Sample	Wave asked in
Census ethnic group[a]	w_racel, w_ racelo_code	2011 UK Census ethnic group	All	First interviewed
	w_racelt, w_ racelwt, w_ racelmt, w_ racelat, w_ racelbt, w_ racelot_code	2011 UK Census ethnic group (telephone version)	All	First interviewed (wave two onwards)
	racel_bh	2001 UK Census ethnic group	BHPS	Asked in BHPS before 2009
	race_bh	1991 UK Census ethnic group	BHPS	Asked in BHPS before 2009
	w_racel_dv	Combined ethnic group	All	

		(derived)		
Religion	w_oprlg1, w_nirel	Religion belong to	All	Repeated. First asked wave 1. See Table 2 for rotation.
	w_oprlg0ni, w_oprlg0	Religion brought up in (asked if does not belong to a religion)	All	

Table 8: Key variables for ethnic group measurement and ethnicity analysis (continued)

Question	Variable Name	Desription	Sample	Wave asked in
Own country of birth[b]	w_ukborn	Whether born in England, Scotland, Wales, Northern Ireland or non-UK country	All	First interviewed
	w_plbornc, w_plbornc_all	Which specific country, if born outside UK	All	
Parents' and grandparents' countries of birth[b]	w_macob, w_macob_all	Mother's country or birth	All	Wave one
	w_pacob, w_pacob_all	Father's country of birth	All	
	w_mgmrob, w_mgmrob_all	Maternal grandmother's country of birth	All	
	w_mgprob, w_mgprob_all	Maternal grandfather's country of birth	All	
	w_pgmrob, w_pgmrob_all	Paternal grandmother's country of birth	All	
	w_pgprob, w_pgprob_all	Paternal grandfather's country of birth	All	
Parent's Ethnic group[c]	w_maid	Mother's ethnic group	All	Wave 1
	w_paid	Father's ethnic group	All	
Strength of identification with parents' ethnic group	w_smaid	Strength of identification with mother's ethnic group, if different from father's ethnic group	If w_racel is not White majority[e]	Wave 1
	w_spaid	Strength of identification with father's ethnic group		
Language/Mother tongue	w_kidlang	The main language that was spoken at home during childhood	All	Wave 2
Ethnic identity	w_ethid*	Whether own country, own language, parents cob if different from own etc are "important to your sense of who you are"	Extra 5 minutes	Repeated. First asked wave two. See Table 2 for rotation
	w_ethclose*	Whether feel happy to meet someone from same country as own etc.	Extra 5 minutes	
	w_pride*	Pride in own country of birth etc.	Extra 5 minutes	
National identity	w_natid1-w_natid6, w_natid97	Whether national identity is English, Scottish, Welsh, Northern Irish, British, Irish, Other (code all that apply)	All	First interviewed

Britishness	w_britid	How important is being British to the respondent	Extra five minutes	Repeated. First asked wave 1. See Table 2 for rotation.

Table 8: Key variables for ethnic group measurement and ethnicity analysis (continued)

Question	Variable Name	Description	Sample covered	Wave asked in
Ethnic minority boost sample screening question[d]		Whether anyone living at the screened address comes from, or have parents or grandparents from any of the listed ethnic minority groups	Screened households	Prior to Wave 1 (fieldwork)
Ethnicity of 3 best friends (self-completion)	w_netet	Ethnic group of best friend/s (UK 2001 Census Ethnic group version)	All	Wave 3
Ethnicity of friends	w_simrace	Proportion of friends that are of the same ethnic group as respondent	All	Wave 3

Notes:
[a]BHPS sample members were asked in wave 18, the harmonized version is available in xwavedat file
[b]For all country of brith variables, those who report other country are asked to name those countries and these have been coded later. Variables that include the names of the other countries have a suffix _all and are available in the Special License version of the data
[c]Coding frame is different from Census Ethnic Group question (w_racel)
[d]Coding frame is different from Census Ethnic Group question (w_racel). The exact question and the response options are included in Address Record Form (ARF (EMB)).
[e]Defined as those who choose White – British/English/Scottish/Welsh/Northern Irish as their ethnic group

For those carrying out analysis of those answer the Extra Five Minutes, the extra five minute sample can be identified by the flag w_xtra5min. We discuss the related issues of weighting for the different samples next.

The EMB sample, and the GP samples can be distinguished using the flag w_hhorig or hhorig. However, analysis should not be carried out separately on the EMB, since the weights are not estimated for that sample specifically.

7. ISSUES FOR ANALYSIS: SAMPLE DESIGN, NON-RESPONSE, WEIGHTING

The complex sample design of Understanding Society and possibly non-random non-response has serious consequences for analysis.

7.1 WHY USE WEIGHTS?

Firstly, the sample design is such that some sections of the population are selected with higher probability than others. Secondly, not everyone eligible for interview are interviewed. Some cannot be contacted, and some of those contacted refuse to give an interview. Finally, if those selected with higher selection probability are different in terms of some variable of interest from those selected with lower probability, then population estimates based on sample statistics will be biased. Similarly if respondents are systematically different from non-respondents in terms of the variable of interest. In such cases (appropriately) weighted estimates provide unbiased population estimates.

For example, Table 9 provides a notional example where the population consists of two groups: 1000 individuals from group A and 500 individuals from group B. There are two sample designs. The first one is such that the sampling fraction is the same across two sub-populations: 0.10. This results in a sample of size 150, with 100 individuals from group A and 50 from group A. But in order to analyse group B we require a larger sample. So, a second option is the second sample design where the sample size for both groups is the same, 100. This means that the sampling fractions are different: 0.10 and 0.20.

Table 9: Illustration of impact of not accounting for differential selection probabilities

	Group A	**Group B**
Population size	1000	500
Earnings	30	15
Sample1		
Sample size	100	50
Sampling fraction	0.10	0.10
Weights	10	10
Sample 2		
Sample size	100	100
Sampling fraction	0.10	0.20
Weights	10	5

If the mean earnings of Group A is 30 and that of Group B is 15, then the population mean earnings is = $\frac{(1000\ X\ 30)+(500\ X\ 15)}{1500}$ = $\frac{37500}{1500}$ = 25

Similarly, the sample mean of sample 1 is 25, but that of sample 2 is 22.5. In other words unequal selection probability along with systematic difference in earnings across the two groups results in a biased estimate of the mean population earnings based on sample mean of sample 2.

Weights are computed as inverse of the selection probabilities and so weighted sample mean for sample 2 = $\frac{(100X30X10)+(500\ X\ 15\ X\ 5)}{(100\ X\ 10)+(100\ X\ 5)}$ = 25

These are differences resulting from the design of the sample. There are also differences in response, both at wave 1 and over time that can affect the representativeness of the sample. These also need to be adjusted for by appropriate weights and the available weights include the combined adjustment for both design and non-response / attrition.

To illustrate the role of weights in producing population estimates the weighted and unweighted estimates of ethnic group distribution of adults living in England are shown in Table 10. The 2011 UK Census distribution is also shown. As is clear, in most cases the weighted estimates are closer to the Census distribution than the unweighted estimates.

Table 10: Ethnic group of residents of England

	UK 2011 Census	Understanding Society Wave 1, 2009-10	
		weighted	unweighted
White			
English/Welsh/Scottish/Northern Irish/British	81.1%	85.6%	72.7%
Irish	1.1%	1.4%	0.8%
Gypsy or Irish Traveller	0.1%	0.0%	0.0%
Other White	4.8%	3.6%	3.1%
Mixed/multiple ethnic group			
White and Black Caribbean	0.5%	0.3%	0.8%
White and Black African	0.2%	0.1%	0.3%
White and Asian	0.4%	0.2%	0.4%
Other Mixed	0.4%	0.2%	0.4%
Asian/Asian British			
Indian	2.6%	2.4%	4.6%
Pakistani	1.7%	1.3%	3.5%
Bangladeshi	0.7%	0.5%	2.8%
Chinese	0.8%	0.4%	0.7%
Other Asian	1.5%	0.7%	1.3%
Black/African/Caribbean/Black British			
African	1.6%	1.2%	3.4%
Caribbean	1.1%	0.8%	2.8%
Other Black	0.4%	0.1%	0.2%
Other ethnic group			
Arab	0.4%	0.2%	0.4%
Any other ethnic group	0.6%	0.9%	1.7%

7.2 AVAILABLE WEIGHTS

The weighting strategy is described in Lynn and Kaminska (2010). The underlying logic of weighting is to take account of unequal selection probabilities (e.g., different probabilities of selection in the EMB sample and the NI part of the GP sample) and differential nonresponse by assigning higher weights to those with lower selection probabilities and lower response propensities.

A single weight is not appropriate for all the possible different types of analysis that might be cared out and so different weights are provided with the data for different types of analysis; however, this may make the selection of the appropriate weight seem difficult.

But it is fully explained in the section on weighting adjustments in the User Guide for the main survey. ***It is very important that users consult that section of the User Guide before commencing analysis.***

The section describes the different types of weight variables available, their naming convention and also guidance on which weights to use. Weights vary by the sample being anlaysed (GPS, EMB, BHPS), the units of analysis (adults, youths, households) and survey instruments used (household grid, household questionnaire, adult questionnaire, proxy questionnaire, adult self-completion, extra five minute questions, youth questionnaire). For the complete set of weights see Tables 11 through 16 in the User Manual for the main survey.

In particular see Table 14 which describes the weights to use if you use responses from at least one Extra Five Minutes question. Also, note that even if the analysis is restricted to ethnic minority respondents, researchers should include ethnic minorities across all sample components not just in the EMB sample.

7.3 SAMPLE DESIGN VARIABLES AND ANALYSIS

The design for most components of Understanding Society sample is stratified and clustered. Most statistical softwares assume that the sample design is simple random sample and estimate standard errors of estimators based on that assumption. If these design features are note taken into account it will result in incorrect standard error estimates. Variables which indicate the primary sampling unit (PSU) and the stratum to which the sample member or sample household belongs, are provided with the data. New entrants are assigned the PSU and stratum of the OSM household they have join. The PSU variable is called **psu** and the variable incidicating stratum is **strata**. These are available in **xwavedat**. For ease of use these are also available in wave specific files. In those files the names have a w_ wave prefix although these do not change across waves.

8. APPROACH TO TRANSLATION AND LANGUAGE

One key issue for the ethnicity strand was the extent to which the questionnaire should be translated into other languages and, if so, which languages. Translation of questionnaires is not simply a case of translating word-by-word from an original questionnaire, but ensuring that the construct being measured is equivalent and that responses provide the same information. Much of the work on translating survey instruments has come from the perspective of translating the same instrument in different countries, in order to harmonise the instrument for cross-national comparison. Nevertheless, many of the issues are the same in a multilingual context.

One of the key insights from this literature is that consideration of the need to translate the instrument may have implications for the initial design of the questionnaire and questions. For example, questions will need to be as simple and clear as possible; a question with ambiguous wording will not be a useful question in the original language and is likely to be an even less useful question once translated.

It is important to acknowledge that whilst the language in a question may be quite simple, it is also very specific. The language of the question should be free of jargon or slang, or if these are used then a note for the translators should be written which sets out in more detail the concept the word is being used to convey. At the same time in the translation process there has to be careful attention to the principles behind questions and response categories in the original. Response categories are used as a tool of measurement and so are carefully chosen so as to be neutral, with symmetrical response categories. Any translation procedure will need to ensure that those doing the translation are aware of this.

In terms of the translation process, there are a variety of practices. 'Best practice' is often regarded to be represented by the approach used in the European Social Survey, and their TRAPD; Translation, Review, Adjudication, Pretesting and Documentation (Harkness 2003). This is an iterative process and covers the stages that a questionnaire goes through, often more than once. There are restrictions in the extent to which it is possible to engage in concurrent consideration of questions in different languages and in the possibilty to change questions or suites of questions that are recognised validated measures. However, while much of the literature now suggests that back translation is not an appropriate approach to translation, there is an emphasis on using more than one translator, with the initial translation checked by a second translator and adjudication taken place in the case of disagreements. It is also recognised that a translation pre-test or pilot is an important element of the quality checking of a translation.

In existing UK surveys at the time that UKHLS wave one was in preparation a range of practice was used, from the exclusion of those who did not speak English to the use of translated advance materials and survey leaflets combined with the use of bilingual interviewers, translators or the use of another (adult) household member to translat 'on the fly'. Pre-translated instruments were used in some cases and administered by a bilingual interviewer, or translator, but this was less common.

Given that at the time of preparation of UKLHS the Census had not included a question on language spoken and fluency in English, information on relative English language fluency across different language groups had to be extropolated from other surveys.

On this basis UKLHS developed translated instruments using independent translator and checker and adjudication. The translated instruments were implemented in computer assisted personal interviewing (CAPI) software, obviating the need for separate, paper-based versions of the translation, which are associated with much longer interview times. This meant that bilingual interviewers could simply switch to the language of choice (or between that and English), while non-bilingual interviewers could work alongside a translator, helping them to enter the response directly into the CAPI laptop.

A translation pilot was carried out on selected languages and found that this process largely worked well.

The selection of languages was based, as noted, on inferences from other surveys combined with the expected composition of the UKLHS overall sample. Note that the inferences were based not just on language use but on inferences about English language fluency for speakers of those languages. Hence it was expected for

example, that Hindi speakers would tend to be relatively fluent in English, while Somali speakers were expected to have much lower levels of English language fluency. On that basis the chosen languages for translation were Welsh (required under the Welsh Language Act), Urdu, Punjabi (in both Urdu and Gurmukhi scripts), Gujerati, Bengali, Cantonese, Somali and Arabic.

For all these languages, interviews were carried out using the translation by either accredited bilingual interviewers or by accredited translators alongside the interviewer.

Where respondents spoke languages not available in these translations, practice resorted to the use of bilingual interviewers or translators or household members, translating 'on the fly', for the respondent.

In fact, use of translated interviews was in fact rather low, which is consistent with the findings from other surveys.In wave 1, out of 50,994 interviews, 456 translated individual (adult) questionnaires were used (i.e., around 0.8%).

9.HELP AND SUPPORT

For further help, make use the the Help pages for Understanding Society. These also provide information on Frequently Asked Questions and the opportunity to ask your own questions at the User Forum, which will be sent to those best placed to respond to them.

A series of training courses are offered to introduce new users to the data and deal with key issues of data manipulation and analysis. It is expected that these will include a specialist ethnicity element, following the development and release of the new Immigrant and Ethnic Minority Boost sample that is currently being recruited to enhance the long term analytical potential of the study.

10.CITATION INFORMATION AND CREDITS

Any publication, whether printed, electronic or broadcast, based wholly or in part on the *Understanding Society* data collection provided by the UK Data Service must be accompanied by the correct citation of the data, as provided by an appropriate bibliographic citation. For example, the format for bibliographic references for the first four waves of Understanding Society is as follows:

The University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society*: Wave 1-4, 2009-2013 [computer file]. 6th Edition. Colchester, Essex: UK Data Service [distributor], December 2014. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-6.

All publications must also acknowledge the Institute for Social and Economic Research as the data provider and the UK Data Service as the data distributor. The acknowledgement, which gives credit to sponsors or distributors, is not a replacement for a proper citation of the data.

CITATION OF THIS GUIDE

This ethnicity user guide should be cited as:
McFall, S., Nandi, A. and Platt, L. (2014). Understanding Society*: UK Household Longitudinal Study: User Guide to Ethnicity Research*. Colchester: University of Essex.

This constitues the *first edition* of this Ethnicity Guide. Subsequent updates will be provided as additional relevant information is identified. In particular there will be a new edition to cover the additional issueS presented by the addition of a new Immigrant and Ethnic Minority Boost (IEMB) sample, which will be added to the existing sample from wave 6.

11. ADDITIONAL LINKS

Here are some useful links for information on resources for ethnicity research:

- Afkhami 2002 provides a useful introduction to ethnicity research in the UK
- The ESRC funded Centre on Dynamics of Ethnicity (CoDE) is conducts research on ethnicity and inequality. For more information see http://www.ethnicity.ac.uk/
- **The Fourth National Survey of Ethnic Minorities (FNSEM)** is the fourth in a series of repeated cross-sectional surveys of ethnic minority individuals carried out approximately every 10 years. The FNSEM incldued a sample of around 5000 ethnic minority households and a comparison sample of around 2900 white households. DOI: 10.5255/UKDA-SN-3685-1. Or see here.
- Home Office **Citizenship Survey** (2001, 2003, 2005) and then renamed to Citizenship Survey (2007/08, 2008/09, 2009/10, 2010/11) are repeated cross-sectional individual (not household) surveys in England and Wales. These surveys included a sample of 15,000+ adults with an ethnic minority boost sample of 5,000-6,000 adults. See here.
- **Millenium Cohort Study** in longitudinal cohort study which started with a sample of N children, drawn from those born between September 2000 and August 2001 in England and Wales and from those born between November 2000 and January 2002 in Scotland and Northern Ireland. Children born in these periods in Scotland, Wales and Northern Ireland were over-sampled as were children from disadvantaged areas (in all four countries) and high ethnic minority concentration areas in England. The first interview with the main carer were conducted during 2001/03. Since then interviews have been conducted every few years and the children themselves were interviewed for the first time in 2014. See here.
- **Labour Force Survey** is a survey of adults living in the UK and has a rotating panel design. There is no ethnic minority boost sample but given its large sample size of around 57,000+ households, the sample of ethnic minority sample members is relatively large, and pooled waves of the study have been extensively used for ethnicity and immigration research. See here.
- **British General Election Studies 1997 – Ethnic Minority Survey** and the **British Election Study Ethnic Minority Survey, 2010 (EMBES)** are part of a

series of election surveys. These surveys are repeated cross-sectional individual surveys which collect information to understand change in political beliefs and voting patterns. In 1997 and 2010 supplementary surveys to the main British Election Studies aimed to cover ethnic minorities respondents were carried out. The 2010 EMBES can be found here.

- **Health Survey for England** is a repeated cross-sectional household surveys which started in 1991 and included ethnic minority boost samples from 1999 to 2004. See here
- **Crime Survey of England and Wales** is a repeated cross-sectional survey of adults and young people which included ethnic minority boost samples for certain years (1988-1996, 2000-2006/07). See here.
- The **ONS Longitudinal Study** is a one per cent sample of the 1971 England and Wales census that has been tracked over subsequent decennial censuses and has been refreshed with those new births and immigrants matching the original sampling criteria. While it does not have an oversample of ethnic minorities, the fact that it represents one per cent of the population of England and Wales provides large minority samples for anlaysis. See further the user support team, CeLSIUS.
- And of course the decennial **UK Censuses**.
- For a detailed list of UK surveys which include ethnicity questions and the sample sizes of ethnic minority sample members see http://www.esds.ac.uk/government/docs/ethnicityintro.pdf

12. REFERENCES

Berthoud, R., Fumagalli, L., Lynn, P. and Platt, L. (2009) *Design of the Understanding Society ethnic minority boost sample*. Understanding Society Working Paper 2009-02. Colchester: University of Essex.

Burton, J., Nandi, A. and Platt, L. (2010) 'Measuring ethnicity: challenges and opportunities for survey research', *Ethnic & Racial Studies* 33 (8): 1332 – 1349.

Harkness, J. A. 2003. 'Questionnaire Translation.' Pp. 35-56 in *Cross-Cultural Survey Methods*, edited by Harkness, J. A., F. J. R. van de Vijver, and P. Ph. Mohler. Hoboken, NJ: Wiley.

Knies, Gundi (ed.) (2014). Understanding Society –*UK Household Longitudinal Study: Wave 1-4, 2009-2013, User Manual*. Colchester: University of Essex.

Lynn, P. (2009) *Sample design for Understanding Society*. Understanding Society Working Paper 2009-01. Colchester: University of Essex.

Lynn, P., Burton, J., Kaminska, O., Knies, G. and Nandi, A. (2012) *An initial look at non-response and attrition in Understanding Society*. Understanding Society Working Paper 2012-02. Colchester: University of Essex.

13. ETHNICITY AND MIGRATION RESEARCH USING UNDERSTANDING SOCIETY

JOURNAL ARTICLES AND WORKING PAPERS

Knies, G., Nandi, A., Platt, L. (2014) Life satisfaction, ethnicity and neighbourhoods: is there an effect of neighbourhood ethnic composition on life satisfaction? *ISER Working Paper* No. 2014-08

Longhi, S., (2014) Cultural diversity and subjective well-being, *IZA Journal of Migration*, **3**:13

McAloney, K., Graham, H., Law, C., Platt, L., Wardle, H., Hall, J. (2014) Fruit and vegetable consumption and sports participation among UK Youth*, International Journal of Public Health*, 59: 117-121

McAloney, K. (2014) 'Mixed' Religion relationships and well-being in Northern Ireland, *Journal of Religion and Health*, 53 (4): 1036-1045

Morando, G. (2014) Partner ethnicity and ethnic minority socio- economic occupation: evidence from the UK, *ISER Working Paper* 2014-29.

Nandi, A. and Platt, L. (2014) Britishness and identity assimilation among the UK's minority and majority ethnic groups, *ISER Working Paper* 2014-01.

Nandi, A. and Platt, L. (2014) A note on ethnicity and identity among the UK born population in Understanding Society, *ISER Working Paper* 2014-04.

Tippett, N., Wolke, D. and Platt, L. (2013) Ethnicity and bullying involvement in a national UK youth sample, *Journal of Adolescence* 36: 639-649.

Uskul, A. and Platt, L. (2014) A note on maintenance of ethnic origin diet and healthy eating in Understanding Society, *ISER Working Paper* 2014-03.

ZuccottiI, C. V. (2014). Do Parents Matter? Revisiting Ethnic Penalties in Occupation among Second Generation Ethnic Minorities in England and Wales. *Sociology*. DOI:10.1177/0038038514540373

BOOK CHAPTERS AND REPORTS

Brown, L., Heath, A., Ji, Yaojun, Nazroo, J. (2013) Addressing Ethnic Inequalities in Social Mobility, *Research findings from the CoDE and Cumberland Lodge Policy Workshop*

Kaufman, E. and Harris, G. (2014*) Changing Places*. London: Demos.

Kaufman, E. and Harris, G. (2014) Demystifying white flight in *Understanding Society: Insights 2014*. Colchester: ISER, University of Essex.

Li, Y. and Heath, A. (2014) Inheriting success in *Understanding Society: Insights 2014*. Colchester: ISER, University of Essex.

Nandi, A. and Platt, L. (2012) How diverse is the UK? in S. McFall (Ed) *Understanding Society: Findings 2012*

Nandi, A. and Khan, O. (2012) Understanding Remittances in Understanding Society: Quantifying Ties Overseas and Ties to Britain, in S. McFall (Ed) *Understanding Society: Findings 2012*

Nandi, A. And Saggar, S. (2012) Employment and perceived racial discrimination, in S. McFall (Ed) *Understanding Society: Findings 2012*

Platt, L. (2014) Ethnic minority inequalities in the job market in *Understanding Society: Insights 2014*. Colchester: ISER, University of Essex.

Please let us know about any research you are doing using *Understanding Society* data!!